Doubly Contrastive End-to-End Semantic Segmentation for Autonomous Driving under Adverse Weather (Supplementary)

Jongoh Jeong
jojeong@rit.kaist.ac.kr
Jong-Hwan Kim
johkim@rit.kaist.ac.kr
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, Republic of Korea

Supplementary Materials

1 Pixel-wise Supervised Contrast

Segmentation-Aware Hard Anchor Sampling. Hard negative sampling is key to contrastive learning [1, 2, 3] as many contrastive methods have shown that various hard negative sampling strategies have, in fact, been useful in discriminating features of the same and different labels. However, it is still up to question whether such methods are practical. Increasing the probability of encountering *hard* negatives leads to a large batch size requirement, and having a memory bank storing exemplar or all (encoded) data features is time- and memory-consuming in the long run. In this light, we treat the task of sampling hard examples as sampling anchors that fail to categorize the pixel correctly in a given image by adopting the anchor sampling strategy, except the memory bank, as proposed in [4] in order to avoid resource overheads in time and memory.

Inter-Image *vs.* Intra-Image Contrast. Inter- and intra-image pixel contrast each refers to comparing pixels in each image and across all images, respectively. It is intuitive to compare each pixel against those in the rest of the images for global consistency and is verified by [4]. Although intra-image contrast (combined with the CE loss) leads to some improvement in mIoU, inter-class method achieves far significant performance gain.

In our pixel contrast loss term, however, we seek to avoid using a large memory banks holding a small number of randomly selected pixel-wise embeddings from each mini-batch as in [4] due to a practical resource constraint. Therefore, we instead use intra-image pixel contrast for a given image (in a mini-batch) and complement it with weak supervision using image-level annotations (*e.g.*, weather conditions). This has an effect of taking into account pixels in each image as well as weather-specific images by its class.

2 Detailed Results

Table 1: Category-specific Semantic Segmentation Performance by Adverse Weather Conditions using SwiftNet (detailed results for the Table 2 in the manuscript). The experiment number follows those in the manuscript.

Bb	Exp.	Cond.	mIoU	Road	Sidewalk	Building	Wall	Fence	Pole	Traffic Light	Traffic Sign	Vege-tation	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	Motor-cycle	Bicycle
ResNet-18	(a)	All	62.63	82.03	73.52	84.46	50.26	41.77	58.72	70.93	59.99	84.24	<u>48.13</u>	94.60	**51.23**	19.58	**85.51**	**54.27**	74.95	85.60	28.11	41.98
		Fog	68.95	91.96	81.47	88.78	67.67	40.79	64.50	74.00	69.32	90.10	59.53	97.94	54.62	47.94	86.69	61.40	80.06	75.76	25.52	52.10
		Night	45.71	73.35	71.50	76.41	37.25	40.45	53.29	59.58	42.45	63.06	11.96	78.47	39.15	4.19	68.99	6.78	0.00	89.32	13.22	39.05
		Rain	63.32	86.12	69.91	87.67	39.65	36.966	56.86	68.93	70.43	73.75	61.08	98.22	64.53	20.48	87.13	42.57	81.86	63.26	26.80	48.59
		Snow	65.43	87.12	73.65	87.95	62.96	49.69	59.68	76.92	69.24	89.16	54.76	970.62	50.18	0.00	87.71	53.48	61.30	88.79	44.85	48.42
	(b)	All	64.04	93.84	76.04	84.73	49.39	42.55	<u>60.21</u>	70.96	61.03	<u>85.61</u>	45.88	**95.34**	<u>50.03</u>	20.31	<u>84.74</u>	51.47	**81.30**	<u>86.38</u>	31.24	<u>45.67</u>
		Fog	71.59	96.81	84.39	88.51	68.95	44.39	67.89	74.52	68.50	90.12	57.91	98.20	53.00	58.31	86.15	60.42	90.58	88.16	35.61	47.69
		Night	47.84	93.88	72.77	77.24	35.34	39.37	53.32	56.90	45.78	69.38	9.69	82.27	38.07	15.16	63.53	3.17	0.00	87.88	19.54	45.67
		Rain	63.90	91.86	74.36	87.54	37.71	40.88	58.87	71.72	68.78	92.27	60.08	98.36	62.04	0.64	87.48	40.39	85.06	71.90	36.43	47.80
		Snow	65.57	92.58	76.40	88.34	63.89	48.15	59.26	74.38	69.93	89.30	52.27	97.77	49.38	0.00	87.47	51.92	72.53	87.15	41.78	43.26
	(f)	All	65.07	94.52	<u>77.03</u>	<u>84.97</u>	<u>54.10</u>	**45.65**	60.11	<u>72.27</u>	**61.68**	85.51	47.93	<u>95.02</u>	49.44	<u>25.59</u>	83.14	48.89	78.77	**86.76**	**39.96**	45.08
		Fog	72.45	96.97	85.07	88.75	70.11	48.28	68.35	74.47	69.93	90.19	58.80	98.14	50.91	60.81	84.56	55.10	93.09	85.65	45.04	52.33
		Night	48.57	94.25	73.53	77.57	44.19	41.27	53.11	61.14	46.03	68.51	11.45	81.20	34.80	22.30	59.45	5.36	0.00	88.45	16.63	43.52
		Rain	63.95	93.06	74.81	88.84	38.16	44.76	57.57	71.84	68.37	92.32	61.35	98.28	57.92	0.00	86.08	44.94	84.32	73.35	37.83	41.20
		Snow	68.31	93.85	78.43	87.50	64.51	52.71	60.58	76.71	71.15	89.53	56.16	97.69	58.59	3.24	86.76	46.15	63.11	87.68	68.82	54.66
	(g)	All	65.38	**94.62**	**77.55**	**85.11**	**55.04**	<u>44.51</u>	**60.51**	**72.35**	<u>60.78</u>	**85.87**	**48.86**	<u>95.02</u>	49.22	**26.53**	83.12	<u>53.52</u>	<u>79.09</u>	83.97	<u>37.26</u>	**49.29**
		Fog	67.94	97.04	85.11	89.00	70.71	46.75	67.27	73.69	66.68	90.07	57.65	98.07	46.94	48.49	85.39	62.01	83.91	62.28	26.69	52.19
		Night	48.56	94.67	74.70	77.76	45.55	40.58	54.84	60.55	44.25	68.93	13.65	81.30	29.05	19.94	60.67	2.85	0.00	88.38	16.95	48.04
		Rain	65.38	92.88	75.13	88.66	37.72	41.24	59.28	72.59	70.17	92.39	62.53	98.30	65.39	21.01	84.82	38.23	82.93	64.57	45.81	48.67
		Snow	68.64	93.72	78.41	88.26	66.08	54.05	59.67	77.26	70.30	89.73	54.68	97.74	56.84	0.00	86.73	56.04	71.08	89.39	58.16	55.97

References

[1] Yuki Markus Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. *arXiv preprint arXiv:1911.05371*, 2019.

[2] Maxime Bucher, Stéphane Herbin, and Frédéric Jurie. Hard negative mining for metric learning based zero-shot classification. In *European Conference on Computer Vision*, pages 524–531. Springer, 2016.

[3] Yannis Kalantidis, Mert Bulent Sariyildiz, Noe Pion, Philippe Weinzaepfel, and Diane Larlus. Hard negative mixing for contrastive learning. *Advances in Neural Information Processing Systems*, 33:21798–21809, 2020.

[4] Wenguan Wang, Tianfei Zhou, Fisher Yu, Jifeng Dai, Ender Konukoglu, and Luc Van Gool. Exploring cross-image pixel contrast for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7303–7313, 2021.